

Parameterizing Region Covariance: An Efficient Way To Apply Sparse Codes On Second Order Statistics

Xiyang Dai
University of Maryland
xdai@cs.umd.edu

Sameh Khamis
University of Maryland
sameh@umiacs.umd.edu

Yangmuzi Zhang
University of Maryland
ymzhang@umiacs.umd.edu

Larry S. Davis
University of Maryland
lsd@umiacs.umd.edu

February 10, 2016

Abstract

Sparse representations have been successfully applied to signal processing, computer vision and machine learning. Currently there is a trend to learn sparse models directly on structure data, such as region covariance. However, such methods when combined with region covariance often require complex computation. We present an approach to transform a structured sparse model learning problem to a traditional vectorized sparse modeling problem by constructing a Euclidean space representation for region covariance matrices. Our new representation has multiple advantages. Experiments on several vision tasks demonstrate competitive performance with the state-of-the-art methods.

1 Introduction

Sparse representations have been successfully applied to many tasks in signal processing, computer vision and machine learning. Many algorithms[1, 12] have been proposed to learn an over-complete and reconstructive dictionary based on such representations. These algorithms involve vectorizing the input data which can destroy inherent ordering information in the data[9, 31]. Instead sparse codes can be constructed directly based on the original structure of the input data. Such structures in-

clude diffusion tensors, region covariance, etc. The region covariance structure, introduced by Tuzel *et al.*[33] provides a natural way to fuse different features for a given region. Additionally, the averaging filter in covariance computation reduces noise that corrupts individual samples. Furthermore, Porikli *et al.*[28] showed that it can be constructed for arbitrary-sized windows in constant time using integral images. Hence, it has become a popular descriptor for face recognition[24, 11, 39], human detection[34], tracking[34], object detection [8, 32], action recognition [38, 7] and pedestrian detection [35].

However, region covariance matrices are positive definite matrices, forming a connected Riemannian manifold. Current manifold-based methods for region covariance often require complex computation. Many applications remain restricted to k-nearest-neighbors or kernel SVMs, using geodesic distance measurement[26, 34, 35]. Pennec *et al.* [26] first introduced the general framework to calculate the statistics based on an affine-invariant metric. Recently, there have been several attempts to develop sparse coding for region covariance matrices[7, 8, 31, 14, 32, 39]. However, such approaches all involve complex computations, including calculating eigenvalues, matrix logarithms and matrix determinants.

We present an approach for sparse coding parameterized representations of region covariance matrices inspired by finance applications. This representation pre-

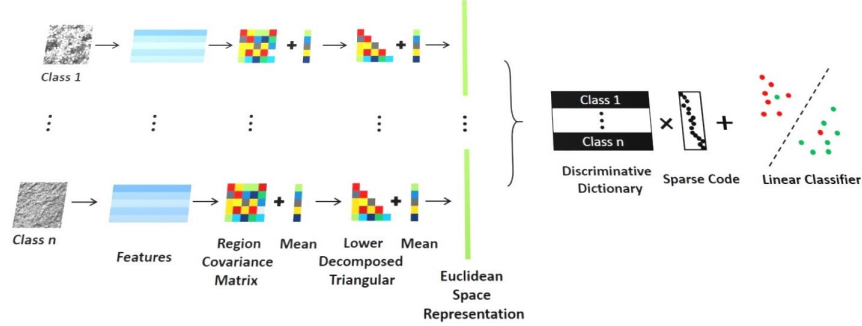


Figure 1: The framework of our approach when applied to texture classification problem.

serves the same second order statistics as region covariance matrices. More importantly, the representation is Euclidean and hence can be vectorized and computed effectively in the traditional sparse coding framework. We further learn discriminative dictionaries over this representation by integrating label consistency regularization and class information into the objective function. The framework of our approach is shown in Figure 1. The main contributions of this paper are:

- Introduction of covariance parameterization used in finance to the computer vision community.
- Design of a new Euclidean representation for region covariance that has multiple advantages, including lower time complexity for measuring similarity and preserving both first order and second order statistics of a given region.
- Performing discriminative dictionary learning on our new representation of region covariance to show its effectiveness.
- Experiments show state-of-the-art performance on multiple tasks.

2 Background

We provide a brief review of the region covariance descriptor and its corresponding similarity measurement methods.

2.1 Region Covariance Descriptors

Given an image I , let Φ be a function that extracts a d -dimension feature vector z_i at each pixel $i \in I$, i.e. $\Phi(I, x_i, y_i) = z_i$, where $z_i \in R^d$ and (x_i, y_i) is the location of pixel i . Φ can be any feature mapping function such as intensity, gradient, different color channels, filter responses, etc. F is a $W \times H \times d$ dimensional feature matrix extracted from I . A given image region R is represented by the $d \times d$ covariance matrix C_R of the set of feature vectors $\{z_i\}_{i=1}^N$ of all N points inside the region R . The region covariance descriptor $C_R \in R^{d \times d}$ is defined as:

$$C_R = \frac{1}{N-1} \sum_{i=1}^N (z_i - \mu_R)(z_i - \mu_R)^T \quad (1)$$

where, μ_R is the mean vector,

$$\mu_R = \frac{1}{|N|} \sum_{i=1}^{|N|} z_i \quad (2)$$

2.2 Positive Definite Similarity Computation

In general, covariance matrices are positive definite, except for some special cases. They are usually regularized to make them strictly positive definite. Hence, the region covariance descriptors belong to the $d \times d$ positive definite space S_{++}^d , which lies on a Riemannian manifold, not in Euclidean space. This fact makes the similarity measurement between two covariance matrices non-trivial. One

well-known method for computing similarity is the Affine Invariant Riemannian Metric (AIRM)[26] which uses the corresponding geodesic distance on the manifold as a similarity measurement:

$$D_R(X, Y) = \|\log(X^{-1/2}YX^{-1/2})\|_F \quad (3)$$

where $\log(\cdot)$ is the matrix logarithm and $\|\cdot\|_F$ is the Frobenius norm. This method is widely used in classification tasks that involve region covariance. However, the requirement of eigenvalue computation makes it very expensive when used in iterative optimization frameworks.

Many methods have been proposed to improve AIRM. One is the Log-Euclidean Riemannian Metric (LERM)[2]:

$$D_{le}(X, Y) = \|\log(X) - \log(Y)\|_F \quad (4)$$

This method maps the positive definite matrices into a flat Riemannian space by taking the logarithm of the matrices so that the Euclidean distance measurement can be used. While the logarithm for each of these matrices can be evaluated offline, computing the matrix logarithm is still expensive.

More recently, LogDet divergence[14] has been investigated:

$$D_{ld}(X, Y) = \text{tr}(XY^{-1}) - \log\det(XY^{-1}) - n \quad (5)$$

where $\log\det(\cdot)$ is the logarithm of a matrix determinant and $\text{tr}(\cdot)$ is the matrix trace. This method was used in several tensor based sparse coding methods[31, 32, 39]. The LogDet divergence reduces computational complexity by replacing the calculation of eigenvalues with determinants. Also, it avoids the explicit manifold embedding and results in a convex MAXDET problem. However, since the computation of matrix determinants each iteration is still roughly $O(d^3)$, where d is the column size of the region covariance matrix, the whole optimization process is still costly.

3 A Euclidean Space Representation for Region Covariance

In this section, we introduce our methods to construct a small set of points that lie in Euclidean space and preserve the second order statistics.

3.1 Understanding the Region Covariance

Covariance matrices used in finance usually represent the variance of stock price and the correlations between different stocks. Region covariance in computer vision applications shares similar concepts. Given a set of features $Z = [f^1, f^2, \dots, f^n]$, for all N points in a region, the region covariance can be written as:

$$\begin{pmatrix} \sum_{i=1}^N (f_i^1 - \mu_{f^1})^2 & \dots & \langle f^1 - \mu_{f^1}, f^n - \mu_{f^n} \rangle \\ \vdots & \ddots & \vdots \\ \langle f^n - \mu_{f^n}, f^1 - \mu_{f^1} \rangle & \dots & \sum_{i=1}^N (f_i^n - \mu_{f^n})^2 \end{pmatrix} \quad (6)$$

where f_i^n is the n th feature value for point i and μ_{f^n} is the mean of the n th feature vector. The diagonal entries of the covariance matrix represent the variances of each feature, while the entries outside the diagonal represent the correlations of different features. To design a covariance representation, we want to include both of these terms.

3.2 Cholesky Decomposition

A meaningful region covariance matrix C should be symmetric and positive semidefinite, and hence can be decomposed as the product:

$$C = LL^T \quad (7)$$

A obvious way to calculate L is using Cholesky decomposition, which enjoys low computation cost and preserves some properties of the covariance matrix[10]. Let L_x, L_y be the lower triangular matrices calculated from C_x, C_y using Cholesky decomposition, the distance between C_x, C_y can be approximated by

$$D_{chol}(C_X, C_Y) = \|L_X e - L_Y e\|_F \quad (8)$$

where e is a standard Euclidean basis. The Cholesky decomposition guarantees that the new representation $s = Le$ is unique for any covariance matrix C .

Although the representation based on Cholesky decomposition works in practice, it is difficult to interpret the entries in the lower triangular matrix. In particular, it is difficult to obtain the correlation coefficients which are available in the original covariance matrix.

3.3 Spherical Decomposition

Alternatively, we seek a lower triangular representation that not only obeys the decomposition rule, but also possesses better statistical interpretations. Inspired by the spherical parametrization method used in finance application[27] for covariance estimation, a new representation can be constructed using spherical coordinates, which involves a series of rotational mappings from the standard basis to the lower triangular matrix[30]. We start with the lower triangular matrix L generated from Cholesky decomposition, and then represent it as:

$$L_{i,j} = \begin{cases} \tilde{s}_{i,1} \cos(\tilde{s}_{i,2}) & j=1 \\ \tilde{s}_{i,1} \cos(\tilde{s}_{i,j+1}) \prod_{k=2}^j \sin(\tilde{s}_{i,k}) & 2 \leq j \leq i-1 \\ \tilde{s}_{i,1} \prod_{k=2}^i \sin(\tilde{s}_{i,k}) & j=i \\ 0 & i+1 \leq j \leq n \end{cases} \quad (9)$$

where $\tilde{s}_{i,j}$ denotes the new representation, $L_{i,j}$ is an element of L . A special case of 9 is $L_{1,1} = \tilde{s}_{1,1}$. To ensure the uniqueness of converting from a covariance matrix to spherical coordinates, we must have:

$$\begin{cases} \tilde{s}_{i,1} > 0, & i = 1, \dots, n \\ \tilde{s}_{i,j} \in (0, \pi) & i = 2, \dots, n \quad j = 2, \dots, i \end{cases} \quad (10)$$

This new representation has the following statistical advantages:

- The diagonal entries of the covariance matrix are captured directly by the entries of this new representation: $C_{i,i} = \tilde{s}_{i,1}^2$.
- Some of the correlation coefficients ρ for the covariance matrix can be uniquely mapped to the new representation: $\rho_{1,i} = \cos(\tilde{s}_{i,2})$, $i = 2, \dots, n$.
- Elements of the new representation are independent of each other.

The new representation lies in Cartesian space [30], hence the distance can be measured using the Frobenius norm:

$$D_{sphere}(C_X, C_Y) = \|\tilde{s}_X - \tilde{s}_Y\|_F \quad (11)$$

where \tilde{s}_X and \tilde{s}_Y are the new representations calculated by 9.

3.4 Combine with the Mean

The mean of the original features can be concatenated to s to make it more informative and robust:

$$s = \lambda \mu_R \cup s \quad (12)$$

λ is a parameter that balances the scale difference between the mean and our representation.

Our representation s lies in Euclidean space and the similarity between representations can be simply measured by the Frobenius norm. Compared to the traditional covariance matrix, our new representation enjoys several advantages:

- Lower time complexity for measuring similarity. The time complexity of using the Frobenius norm to measure the similarity is $O(d^2)$ compared to $O(d^3)$ of AIRM[26] and LogDet[14].
- Informative and robust. Our new representation preserves both the first and the second order statistics. Since the region covariance only captures the differences between features, it may lose some useful statistics within separate feature channels. Hence, fusing feature means into our representation enhances robustness.
- Flexibility. The similarity measurement of our new representation can be calculated in Euclidean space, which enables applying many traditional learning methods to second order statistic.

4 Discriminative Sparse Coding and Dictionary Learning

We next describe a method to learn a reconstructive and discriminative dictionary from multi-class data. We construct a sub-dictionary for each class. We explicitly encourage independence between dictionary atoms from different sub-dictionaries and leverage class information in the optimization problem. We adopt the LC-KSVD[12] method to learn the discriminative dictionary.

4.1 Dictionary Learning via Label Consistent Regularization

Let S be a set of N d -dimensional Euclidean space region covariance representations for training dictionary,

i.e. $S = [s_1, s_2, \dots, s_N] \in R^{d \times N}$. Learning a reconstructive dictionary with K atoms for sparse representation of S can be formulated as:

$$\arg \min_{D, X, A, W} \|S - DX\|_F^2 + \alpha \|Q - AX\|_F^2 + \beta \|H - WX\|_F^2$$

$$s.t. \quad \forall i, \|x_i\|_0 \leq T \quad (13)$$

where $D \in R^{d \times K}$, $K \geq d$ is the learned over-complete dictionary. $X \in R^{K \times N}$ is the sparse codes for given inputs, $Q \in R^{K \times N}$ are the "discriminative" sparse codes, A is a linear transformation matrix defined to transform the original sparse codes X to be most discriminative in sparse feature space, W denotes the parameters of a linear classifier $f(x; W) = Wx$, $H \in R^{m \times N}$ are the class labels and T is a sparsity constraint factor.

Minimizing the objective function not only encourages independence between dictionary atoms from different sub-dictionaries, but also trains a linear classifier simultaneously. We use the efficient K-SVD algorithm to find the optimal solution for all parameters simultaneously.

4.2 Classification

After obtaining the dictionary D and the linear classifier parameter W , the sparse representation X_{test} for the test inputs S_{test} can be calculated as:

$$X_{test} = \arg \min_X \|S_{test} - DX\|_F^2$$

$$s.t. \quad \forall i, \|x_i\|_0 \leq T \quad (14)$$

We simply use the linear classifier $f(x; W) = Wx$ to estimate the label of a test sample $x_i \in X_{test}$:

$$l = \arg \max_l (Wx_i) \quad (15)$$

4.3 Sparse Codes as Features

We can also fuse the generated sparse codes with other features. One drawback of 14 is that the sparsity constraint factor T is a hard threshold that forces the sparse codes to have fewer than T non-zero items. This is good for classification task, but when using sparse codes as features we are not concerned with the number of non-zero items. Instead we want to make the sparse codes as informative as possible. Hence, we consider a "soft" version

of equation 14:

$$X_{test} = \arg \min_X \|S_{test} - DX\|_F^2 + t_1 \|X\|_1 + \frac{t_2}{2} \|X\|_2^2 \quad (16)$$

where t_1 and t_2 are the new sparsity constraint factors. These two parameter control the generation of more continuous sparse codes.

5 Experiments

We evaluate our approach on several different tasks: texture classification, object classification, face recognition, material classification and person re-identification. Sample images for different tasks are shown in Figure 2. For fair comparison, we experiment on the same features as reported by other methods.

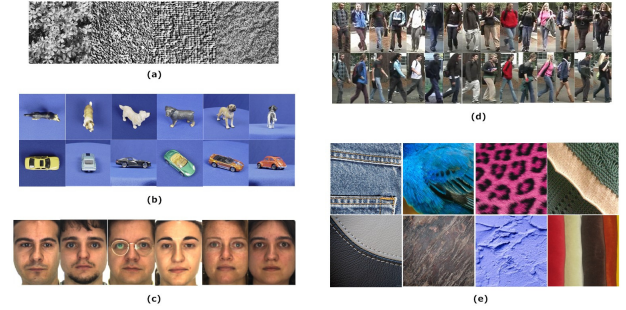


Figure 2: Sample images from experiment datasets for different tasks: (a) Texture classification, (b) Object classification, (c) Face recognition, (d) Person re-identification and (e) Material classification

5.1 Texture Classification

Evaluation Protocol. We follow the protocol in [29] to create mosaics under nine test scenarios. Each scenario has various numbers of classes, including 5-textures, 10-textures and 16-textures. Each image in the dataset is resized to 256×256 and cut into 32×32 non-overlapped blocks, yielding 64 data samples per image. For each scenario, we randomly select 5 data samples as training and use the rest for testing. The evaluation is repeated 10 times.

Scenario	logE-SR[38]	TSC[32]	RSR[8]	SDL[39]	Ours-chol	Ours-sphere
5c	0.88	1.00	0.99	0.99	0.99	0.98
5m	0.54	0.73	0.85	0.95	0.96	0.97
5v	0.73	0.86	0.89	0.90	0.92	0.91
5v2	0.70	0.85	0.89	0.93	0.94	0.95
5v3	0.65	0.83	0.87	0.84	0.97	0.92
10	0.60	0.81	0.85	0.89	0.82	0.81
10v	0.64	0.68	0.86	0.91	0.84	0.79
16c	0.68	0.75	0.83	0.86	0.89	0.89
16v	0.56	0.66	0.77	0.89	0.79	0.79

Table 1: Texture classification results on the Brodatz dataset.

Implementation Details. We extract features $F(x) = \{I, |I_x|, |I_y|, |I_{xx}|, |I_{yy}|\}$ based on intensity and gradient from each sample. They form a 5×5 region covariance matrix and result in a 20-dimension vector in our representation. We use the same parameter configuration ($\sqrt{\alpha} = 5, \sqrt{\beta} = 5, T = 1$) in all test scenarios.

Results. Table 1 shows the classification results under the nine scenarios. We compare our method with logE-SR[38], TSC[32], RSR[8], SDL[39]. The mean accuracy of our method achieves the best result in over half of the scenarios (5m, 5v, 5v2, 5v3, 16c). Overall, our maximum classification results over 10 runs are comparable to the best scores.

Method	Accuracy
TSC[32]	37.1
K-LE-SC[18]	76.6
RSR[8]	81.6
Riem-SC[3]	77.9
Ours-chol	79.8
Ours-sphere	84.0

Table 2: Object classification results on the ETH80 dataset.

shown in Table 2. Our results are comparable to other methods.

5.2 Object Classification

Evaluation Protocol. The ETH80 dataset[17] contains eight objects with ten instances each collected from 41 different views. There are 3280 images total. Images for each object have large view point changes which make this dataset very challenging for object recognition task.

Implementation Details. For each image, we generate a 19×19 covariance matrix with feature $F = \{x, y, R, G, B, |I_x|, |I_y|, |I_{LoG}|, \sqrt{I_x^2 + I_y^2}, F_{Laws}\}$, where I_{LoG} is the responses from Laplacian of Gaussian filter, F_{Laws} is the responses from the bank of Laws texture filters [15];

Results. We randomly split 80% of dataset for training and use the rest for testing. We repeat the procedure 10 times and report the average accuracy. We compare our results with several state-of-the-art methods [32, 18, 8, 3],

5.3 Face Recognition

Evaluation Protocol. The AR face dataset[22] contains over 4000 face images captured from 126 individuals. For each individual, there are 26 images separated in two sessions. We follow the protocol used in [39], randomly select 10 subjects to evaluate in our experiment. We repeat the evaluation 20 times.

Implementation Details. Each image is cropped to 27×20 and converted into gray scale. We extract the intensity and the spatial information along with the responses of Gabor filters with 8 orientations $\theta_u = \frac{\pi u}{8}, u \in \{0, 1, 2, \dots, 7\}$ and 5 scales $v \in \{0, 1, \dots, 4\}$: $F(x, y) = \{I, x, y, |G_{0,0}(x, y)|, \dots, |G_{0,4}(x, y)|, \dots, |G_{7,4}(x, y)|\}$ where $G_{u,v}$ is the response of a 2D Gabor wavelet[16]

Method	Accuracy
SD[20]	43.5%
CDL[37]	52.3%
RSR[8]	52.8%
Ours-chol	57.2%
Ours-sphere	56.8%

Table 4: Material classification results on the UIUC material dataset.

defined by:

$$G_{u,v} = \frac{k_v^2}{4\pi^2} \sum_{t,s} e^{-\frac{k_v^2}{8\pi^2}((x-s)^2 + (y-t)^2)} \\ (e^{ik_v(x-t)\cos(\theta_u) + (y-s)\sin(\theta_u)} - e^{-2\pi^2}) \quad (17)$$

where $k_v = \frac{1}{\sqrt{2v-1}}$.

Results. We randomly select 15, 18, 21 images per person to train and use the rest for testing. Our results are compared with [32, 8, 39]. Table 3 contains our results. We achieve significant performance improvements on all three test configurations.

5.4 Material Classification

Evaluation Protocol. The UIUC material dataset[20] contains eighteen categories with twelve images each (mainly belong to bark, fabric, construction materials, outer coat of animals and so on). Images for each category have different scales and are collected in the wild, which make this dataset very difficult. This dataset is considered as one of the state-of-art benchmarks for material classification task. The standard evaluation protocol is to randomly split half of dataset for training and use the rest for testing. We report the average accuracy over 10 repeats.

Implementation Details. For each image, we generate a 155×155 covariance matrix using 128 dimensional SIFT feature and 27 color feature (3×3 raw RGB pixels around the center of SIFT descriptor). We calculate above region covariance matrices over a 12×12 window size with a 4 step size.

Results. We compare our results with several state-of-the-art methods [20, 37, 8], shown in Table 4. Our results are comparable to other methods.

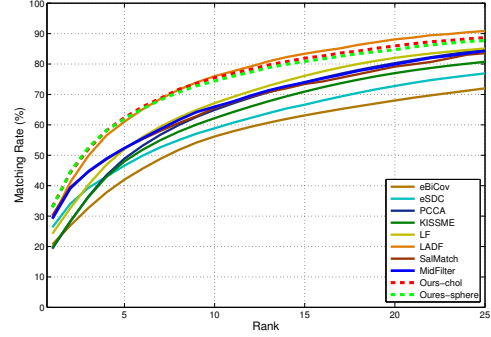


Figure 3: CMC curves on the VIPeR dataset.

5.5 Person Re-identification

Evaluation Protocol. The VIPeR dataset[5] contains 632 pedestrian pairs captured from different camera views. Each image in the pair is resized to 128×48 . They exhibit large viewpoint variations among pedestrian pairs, which makes it one of the most challenging datasets in person re-identification. We follow the protocol widely used in [6], splitting the 632 pedestrian pairs into half for training and half for testing. Two-fold validation is applied during evaluation. We repeat the evaluation 10 times and report the average result.

Implementation Details. We extract 9×9 blocks with a stride of 4 from each image. For each block, we extract gradient and color features in different channels (including RGB, HSV and color name[36]) $F(x) = \{I, |I_x|, |I_y|, R, G, B, H, S, V, cln\}$ to form region covariance matrices. This generates a 10×10 region covariance matrix and result in a 65-dimensional vector in our representation. We then learn sparse codes and use them as features. Additionally, we also extract color histograms from different channels (Lab, HSV and color name[36]) using 7×48 stripes with a stripe of 3 for consistency with our region covariance sparse code feature. The color histograms are further reduced to 300 dimensions by PCA. We concatenate these two features together with normalizing the maximum value to 1 for each sample and use information theoretic metric learning method[4] to learn the final ranks.

Results. We compare our method with state-of-the-art methods that don't require foreground priors such as PCCA[23], KISSME[13], eBiCov[21], eSDC[41],

#train sample	TSC[32]	RSR[8]	SDL[39]	Ours-chol	Ours-sphere
15 per person	78.6	81.4	82.3	86.7	89.1
18 per person	79.9	84.1	85.2	89.0	90.2
21 per person	80.8	85.7	86.1	90.6	91.8

Table 3: Face recognition results on the AR face dataset.

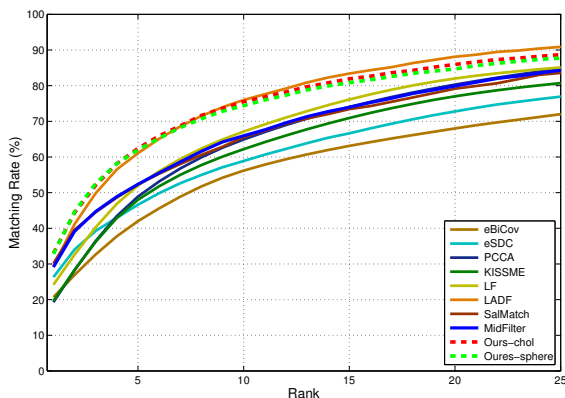


Figure 4: A Example of the matching pairs at each rank on VIPeR dataset. Green box shows the probe images, red box shows the ground truth matches. Each row shows the retrieved results from top matches.

Method	Rank 1 Accuracy
eBiCov[21]	20.66
eSDC[41]	26.31
PCCA[23]	19.27
KISSME[13]	19.60
LF[25]	24.18
LADF[19]	29.34
SalMatch[40]	30.16
MidFilter[42]	29.11
Ours-chol	32.99
Ours-sphere	32.84

Table 5: Rank 1 matching accuracy on the VIPeR dataset.

LF[25], SalMatch[40], LADF[19] and MidFilter[42]. Table 5 shows the rank 1 accuracy on the VIPeR dataset. The rank 1 results of our method outperform all the competing methods. Figure 3 contains the cmc ranking curve from rank 1 to rank 25. Our curve is competitive to most of the state-of-the-arts methods. By visualizing the matching pairs (shown in Figure 4), we find our approach is good at finding discriminative textures thanks to our region covariance representation.

6 Conclusion

We introduced a new representation for region covariance which lies in Euclidean space. This new representation not only shares the same second order statistics with covariance matrices, but also includes the first order statistics. Analysis shows its space and computation advantages over region covariance matrices. Additionally, the discriminative dictionary learning problem on this representation can be solved efficiently in the traditional K-SVD framework. Experiments on different tasks demonstrate the proposed approach is effective and robust.

References

- [1] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *TSP*, 54(11):4311–4322, 2006.
- [2] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56(2):411–421, 2006.
- [3] A. Cherian and S. Sra. Riemannian sparse coding for positive definite matrices. In *ECCV*, 2014.
- [4] J. Davis, B. Kulis, P. Jain, S. Sra, and I. Dhillon. Information-theoretic metric learning. In *ICML*, 2007.

- [5] G. Doug, B. Shane, and T. Hai. Evaluating appearance models for recognition, reacquisition, and tracking. In *PETS*, 2007.
- [6] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *CVPR*, 2010.
- [7] K. Guo, P. Ishwar, and J. Konrad. Action recognition using sparse representation on covariance manifolds of optical flow. In *AVSS*, 2010.
- [8] M. Harandi, C. Sanderson, R. Hartley, and B. Lovell. Sparse coding and dictionary learning for symmetric positive definite matrices: A kernel approach. In *ECCV*, 2012.
- [9] T. Hazan, S. Polak, and A. Shashua. Sparse image coding using a 3d non-negative tensor factorization. In *ICCV*, 2005.
- [10] X. Hong, H. Chang, S. Shan, X. Chen, and W. Gao. Sigma set: A small second order statistical region descriptor. In *CVPR*, pages 1802–1809, June 2009.
- [11] H. Huo and J. Feng. Face recognition via aam and multi-features fusion on riemannian manifolds. In *ACCV*, 2009.
- [12] Z. Jiang, Z. Lin, and L. Davis. Learning a discriminative dictionary for sparse coding via label consistent k-svd. In *CVPR*, 2011.
- [13] M. Kostinger, M. Hirzer, P. Wohlhart, P. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *CVPR*, 2012.
- [14] B. Kulis, M. Sustik, and I. Dhillon. Learning low-rank kernel matrices. In *ICML*, 2006.
- [15] K. Laws. Rapid texture identification. *Proc. SPIE*, 0238:376–381, 1980.
- [16] T. Lee. Image representation using 2d gabor wavelets. *TPAMI*, 18(10):959–971, 1996.
- [17] B. Leibe and B. Schiele. Analyzing appearance and contour based methods for object categorization. In *CVPR*, 2003.
- [18] P. Li, Q. Wang, W. Zuo, and L. Zhang. Log-euclidean kernels for sparse representation and dictionary learning. In *ICCV*, 2013.
- [19] Z. Li, S. Chang, F. Liang, T. Huang, L. Cao, and J. Smith. Learning locally-adaptive decision functions for person verification. In *CVPR*, 2013.
- [20] Z. Liao, J. Rock, Y. Wang, and D. Forsyth. Non-parametric filtering for geometric detail extraction and material representation. In *CVPR*, 2013.
- [21] B. Ma, Y. Su, and F. Jurie. Bicov: a novel image representation for person re-identification and face verification. In *BMVC*, 2012.
- [22] A. Martinez and R. Benavente. The ar face database. In *CVC Technical Report 24*, 1998.
- [23] A. Mignon and F. Jurie. Pcca: A new approach for distance learning from sparse pairwise constraints. In *CVPR*, 2012.
- [24] Y. Pang, Y. Yuan, and X. Li. Gabor-based region covariance matrices for face recognition. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(7):989–993, 2008.
- [25] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian. Local fisher discriminant analysis for pedestrian re-identification. In *CVPR*, 2013.
- [26] X. Pennec, P. Fillard, and N. Ayache. A riemannian framework for tensor computing. *ICJV*, 66(1):41–66, 2006.
- [27] J. Pinheiro and D. Bates. Unconstrained parameterizations for variance-covariance matrices. *Statistics and Computing*, 6:289–296, 1996.
- [28] F. Porikli and O. Tuzel. Fast construction of covariance matrices for arbitrary size image windows. In *ICIP*, 2006.
- [29] T. Randen and J. Husoy. Filtering for texture classification: a comparative study. *TPAMI*, 21(4):291–310, 1999.
- [30] F. Rapisarda, D. Brigo, and F. Mercurio. Parameterizing correlations: a geometric interpretation. *IMA J Management Math*, 18(1):55–73, 2007.
- [31] R. Sivalingam, D. Boley, V. Morellas, and N. Panikolopoulos. Tensor sparse coding for region covariances. In *ECCV*, 2010.
- [32] R. Sivalingam, D. Boley, V. Morellas, and N. Panikolopoulos. Positive definite dictionary learning for region covariances. In *ICCV*, 2011.
- [33] O. Tuzel, F. Porikli, and P. Meer. Region covariance: A fast descriptor for detection and classification. In *ECCV*, 2006.
- [34] O. Tuzel, F. Porikli, and P. Meer. Human detection via classification on riemannian manifolds. In *CVPR*, 2007.
- [35] O. Tuzel, F. Porikli, and P. Meer. Pedestrian detection via classification on riemannian manifolds. *TPAMI*, 30(10):1713–1727, 2008.
- [36] J. van de Weijer, C. Schmid, J. Verbeek, and D. Larlus. Learning color names for real-world applications. *TIP*, 18(7):1512–1523, 2009.
- [37] R. Wang, H. Guo, L. Davis, and Q. Dai. Covariance discriminative learning: A natural and efficient approach to image set classification. In *CVPR*, 2012.
- [38] C. Yuan, W. Hu, X. Li, S. Maybank, and G. Luo. Human action recognition under log-euclidean riemannian metric. In *ACCV*, 2009.
- [39] Y. Zhang, Z. Jiang, and L. Davis. Discriminative tensor sparse coding for image classification. In *BMVC*, 2013.
- [40] R. Zhao, W. Ouyang, and X. Wang. Person re-identification by salience matching. In *ICCV*, 2013.

- [41] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person re-identification. In *CVPR*, 2013.
- [42] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person re-identification. In *CVPR*, 2014.